Optimizing Aromatase Expression and Uncovering Novel Allosteric Inhibitors for Breast Cancer Treatment

A Senior Honors Project Presented to the Faculty of the
Department of Chemistry,
University of Hawai‘i at Mānoa

In Partial Fulfillment of the Requirements
For Bachelor of Science in Biochemistry with Honors

By Michelle Hu
April 15, 2017

Committee:
Dr. Ho Leung Ng, Mentor
Dr. Joseph Jarrett

Acknowledgements

To Dr. Ho Leung Ng and Samson Souza: thank you for taking the time to assist me with my educational endeavors, contributing to my academic growth, and aiding me in the development of this paper with your expertise. To Shalin Zhang, Matt Lam, and Nana Katayama: thank you all for assisting with my experiments and supporting me throughout my senior thesis. To Dr. Joseph Jarrett: thank you for providing critical feedback for my paper. To the Undergraduate Research Opportunities Program: thank you for proving the funding needed to make this project reality. Without all of your help, this project would have not been possible.

Abstract

Breast cancer occurs in 1 of 8 women while 2,600 new cases of breast cancer are expected in men during 2016 alone[1]. Aromatase, a cytochrome P450 enzyme that interconverts androgens into estrogens, is linked to hormonal breast cancer development[2]. Aromatase inhibitors are currently used to treat breast cancer, but the mode of binding for some inhibitors remains unknown. The objective of this project is to optimize aromatase recombinant expression in *E. coli* and discover novel allosteric inhibitors. While screening possible new inhibitory compounds using an activity assay, we identified AR11 and AR13, which produced IC_{50} values of 25.35 μM and 0.41 μM respectively. We have not yet been successful in increasing recombinant expression of mutant-type aromatase, despite adjusting induction time, incubation temperature, and cell strain. Although optimization of aromatase expression was not achieved, possible inhibitors were uncovered which will be used in future screening of protein crystallization conditions once expression is improved. These crystal screens can then be used to generate new structures of aromatase:inhibitor complexes, leading to improved inhibitor potency and reduced toxicity.

Keywords: Aromatase, Aromatase Inhibitors, Inhibition Assays, *E. coli* Expression

Table of Contents

List of Figures

Problem Statement

The development of hormonal breast cancer is linked to estrogen concentrations in the human body[2]. Aromatase is the only enzyme that converts androgens into endogenous estrogens. Scientists are examining the structure of aromatase in order to improve their knowledge of inhibitor binding sites for breast cancer drug development. By doing so, they can create inhibitors with tighter binding for a more personalized treatment of breast cancer.

Current aromatase inhibitors in clinical use may not have the side effects of older drugs such as lethargy, ataxia, and morbilliform skin rash, but they still pose an increased likelihood of musculoskeletal effects such as arthritis, arthralgia, and myalgia[3]. Some people find these side effects so debilitating that they become noncomplacent[4]. Research into aromatase continues in order to create better aromatase inhibitors with less severe side effects. Aromatase is an enzyme encoded by the gene CYP19A1, found on chromosome 15. It functions by catalyzing the oxidation of a methyl group to formate, followed by aromatization of the androgen to estrogen[2,3]. With detailed knowledge of the reaction mechanism, scientists are able to devise possible methods to inhibit the production of estrogen. An example of a clinically useful aromatase inhibitor is letrozole.

Letrozole is a third-generation aromatase inhibitor that is very effective in treating hormonal breast cancer due to high specificity for aromatase, compared to earlier generations of inhibitors[5]. The chemical structures of several aromatase inhibitors, including letrozole, are presented in Figure 1.

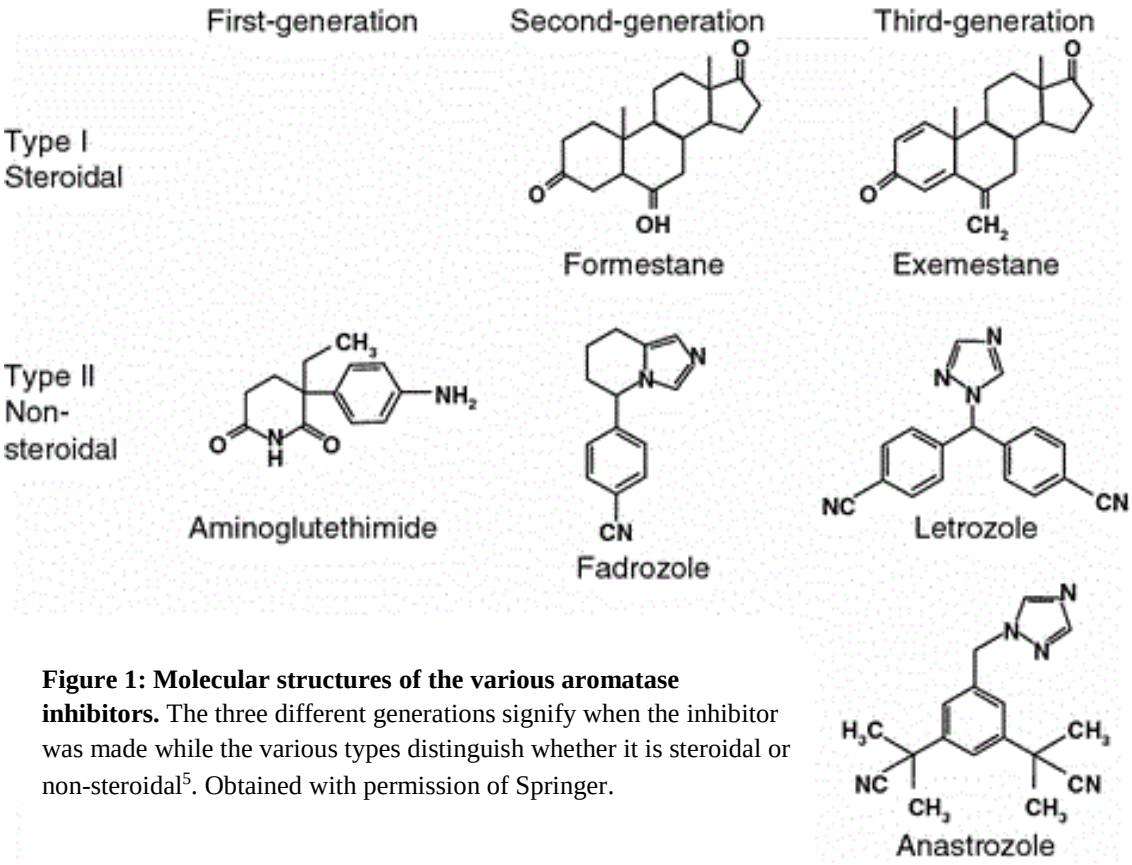

Figure 1: Molecular structures of the various aromatase inhibitors. The three different generations signify when the inhibitor was made while the various types distinguish whether it is steroidal or non-steroidal[5]. Obtained with permission of Springer.

Letrozole's non-steroidal shape causes it to interact at areas outside the active site of aromatase. This is because letrozole's unique shape partially mimics the steroid backbone of androstenedione, the enzyme's substrate[5]. Even though letrozole is the most potent of the inhibitors being used for hormonal breast cancer treatment, there are still some side effects that should be resolved in order to improve its benefits over risks. These side effects include mild symptoms of hot flashes, vaginal dryness, and headaches along with musculoskeletal symptoms mentioned previously due to lack of estrogen[2]. Though it has only been 8 years since aromatase's structure was discovered, there is still much more to be explored. New data regarding aromatase structure will help scientists expand the variety of aromatase inhibitors in clinical use.

Research Objective and Significance

The goal of this project will focus on improving aromatase recombinant expression and finding new aromatase inhibitors. Mutant aromatase enzyme will be expressed following the Lo lab procedure[6] and adjusted until protein expression is optimized. The mutations introduced into the plasmid-encoded gene include replacing 39 amino-terminal amino acids with 10 hydrophilic residues and adding 4 histidine residues at the C-terminus[6]. The final goal will be to crystallize the mutant enzyme with new aromatase inhibitors identified by the Ng lab via activity assays. This project may lead to new inhibitors of aromatase that can be used to treat hormonal breast cancer.

Literature Review

Introduction

New advances in medical research occur every day, especially in the field of breast cancer research. Breast cancer is the leading cause of death for women between ages 30 to 54 and researchers are continuously working towards finding new cures to curb this disease[7]. However, there is still much to be learned about breast cancer as well as the possible methods that can be used to treat it.

Aromatase and Breast Cancer

Aromatase is an enzyme that functions in the production of estrogens by converting androgens through aromatization[2]. Androgens are a broad class of steroids that are associated as male sex hormones. These differ from estrogens which are associated as female sex hormones. Progesterone is a steroid associated with pregnancy and the menstrual cycle. All three hormones are found in humans and are required in varying concentrations based on sex. Cytochrome

$P450_{arom}$ and NADPH-cytochrome P450 reductase form a protein complex that regulate the mechanism seen in Figure 2[8,9]. NADPH-cytochrome P450 reductase acts as an electron source for cytochrome $P450_{arom}$ to aromatize the androgen ring while cytochrome $P450_{arom}$ contains the active site for the reaction to occur[3]. The first two steps involve oxidation using NADPH[9]. The next step is still debated but the consensus seems to be that the ring becomes aromatic and $Fe^{*}O^{3+}$ is used to generate formate at the end[9]. A k_{cat} of 0.06 s^{-1} was reported for aromatase[10]. The small k_{cat} means that the turnover rate of aromatase is slow, suggesting an unfavorable reaction.

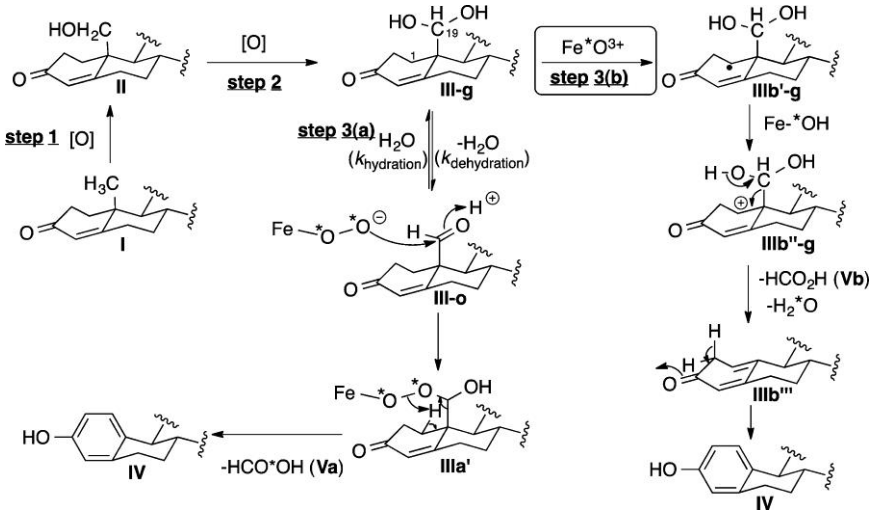

Figure 2: The mechanism of aromatase. Androstenedione, the substrate, is being converted into its product, estrone[9]. Step 3b is the favored mechanism. The figure is available under the terms of the ACS Editors' Choice license.

Various studies investigated the roles of aromatase activity and estrogen levels in breast cancer cells, but it was hard to reach a complete consensus on this relationship because of the variability exhibited between breast cancer cells. Aromatase activity and estrogen concentration have been shown to be inversely related to each other[11]. Estrogen also plays a role in producing growth factors in estrogen responsive tumor cells, being a potent mitogen in estrogen receptor (ER)-α-positive human breast cancer lines, and a key component in tumorigenesis and

progression of most breast cancers[7,12]. Cytochrome P450s have been studied in great detail, uncovering the various types of polymorphisms exhibited by these proteins and helping with the development of alternative cancer drugs that would bind to specific protein variants[13]. While this information gave researchers an understanding of aromatase function, the optimization of drugs that target aromatase would be facilitated by the solution of the three-dimensional structure of the enzyme.

In 2009, aromatase's structure was finally elucidated (Figure 3). Aromatase varies from other cytochrome P450s proteins, as its androgen-specific active site is surrounded by hydrophobic and polar areas to complement the substrate's hydrophobic nature and polar side chains[14]. In order to block this active site, aromatase inhibitors also needed hydrophobic and polar sites to complement the active site. Scientists could now visualize the possible binding sites that would inhibit estrogen production and influence breast tumor growth with the structure of aromatase.

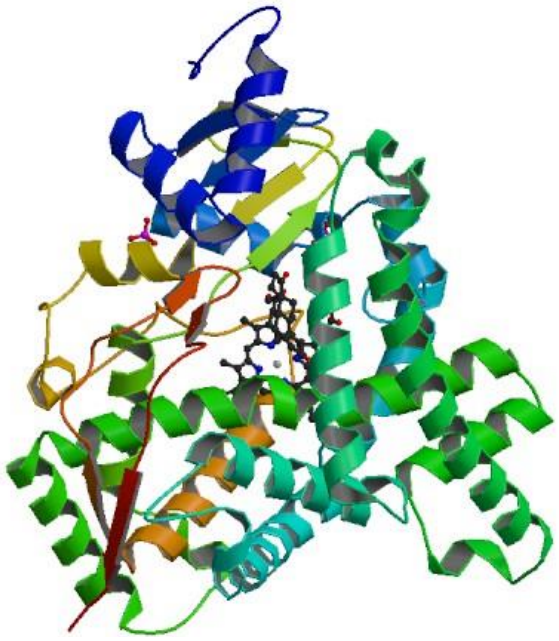

Figure 3: Ribbon depiction of aromatase. The first x-ray crystal structure of aromatase with bound androstenedione in its active site. Black is the heme cofactor, grey is the Fe^{2+} ion[6]. Obtained with permission of Springer.

After discovering the structure, scientists had another hurdle to cross: expressing abundant amounts of stable human aromatase. As mentioned before, aromatase is a hydrophobic molecule with multiple polar side chains[14]. Because of this, it is a very unstable molecule to work with. To make the purification possible, detergents or protein modifications are used to make it stable *in vitro*[15,16]. This is true not only for aromatase but also other P450 enzymes such as CYP1A2 and aldosterone synthase (CYP11B2)[17,18]. Even with the large amounts of research on P450 enzymes, there is still work to be done to optimize expression in *E. coli* as it is the cheaper alternative compared to other model organisms.

Aromatase Drug Development

Aromatase inhibitors are considered the standard of care for hormone-receptor-positive breast cancer in postmenopausal women[4]. Throughout the years, aromatase inhibitors with higher efficacy have been designed to treat breast cancer. However, there are still improvements that can be made to enhance efficacy and reduce toxicity. It is known that hormonal breast cancer depends on estrogen concentration in breast tissue. However, estrogen production has implications in other health problems such as bone loss and nicotine addiction, which should be acknowledged in order to understand the varying effects aromatase inhibitors may have not only in the breast but throughout the body[4,19]. Because of the vast variety of side effects seen with aromatase inhibitors, they can be separated into two types, Types I and II, as well as three generations, first, second, and third generation inhibitors[20].

Type I inhibitors, also called steroidal or orthosteric inhibitors, are a type of molecule that compete directly with androgen for the active site because of its chemical similarities with other steroid molecules[3]. By competing for the active site, inhibitors prevent androgens from being converted to estrogens by aromatase, thus slowing the growth of the breast tumor. However,

once the Type 1 inhibitor is bound to the active site, it cannot be released, inactivating aromatase, hence its other name suicide inhibitors[2]. Type II inhibitors, also called non-steroidal or allosteric inhibitors, are a type of molecule that binds to aromatase through other areas outside the active site[3]. By doing so, they can change the shape of aromatase's active site leading to improper binding between the altered active site and the androgen. Because Type II inhibitors are not structured like other steroid molecules, the effects can be reversed depending on the amount of steroids present in the cell[2]. Though both types of inhibitors bind to aromatase in diverse ways, the main goal remains the same; to reduce aromatase activity, reduce estrogen concentration, and treat breast cancer. Figure 4 shows a diagram of the basic concepts of these processes.

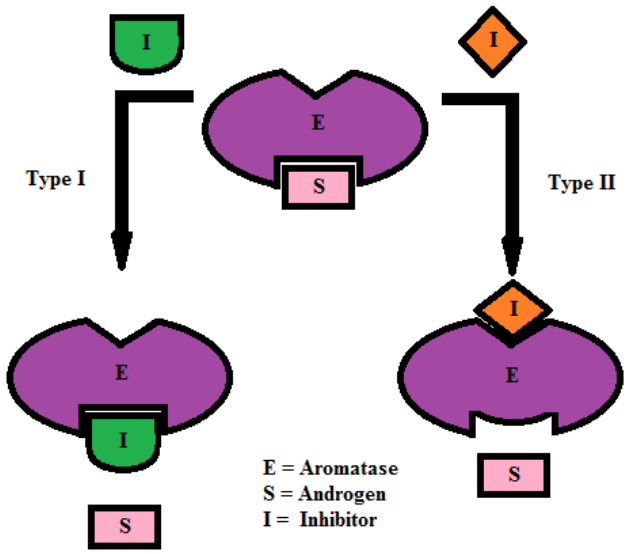

Figure 4: Mechanism for Type I and II inhibitors. The drawn picture shows the typical mechanism when the substrate, androgen, goes to the active site of the enzyme, aromatase and how the enzyme behaves when a Type I and Type II inhibitor is introduced to the reaction. The substrate in both cases cannot bind to the active site.

First, second, and third generation inhibitors are categorized based on when they were developed[20]. The primary first generation drug used to treat breast cancer was aminoglutethimide (AG), a Type II inhibitor. It was discontinued from use because of its low binding specificity to aromatase, leading to drug toxicity due to its interactions with other enzymes in endocrine

systems such as cortisol and aldosterone biosynthesis[2,20]. A new generation of drugs were created to fix this problem named second generation inhibitors. The primary second generation drug in use was 4-hydroxyandrostenedione (formestane), a Type I inhibitor that was more effective than AG but still produced some side effects[2]. Formestane also required intramuscular injection, making it harder to distribute compared to oral tablets[20]. Since their cons outweighed the pros, second generation drugs were eventually phased out in favor of third generation drugs for breast cancer treatment. The third generation cancer drug primarily used now is letrozole, a Type II inhibitor that is considered the most potent breast cancer drug. The reason it is more effective is because of its minimal effects on other endocrine systems in the body, while still inhibiting aromatase present in the peripheral tissue of breast cancer cells[4,5]. Even though aromatase inhibitors are the first line of defense in hormonal breast cancer, there is still the possibility of drug resistance occurring.

Drug resistance occurs when a drug becomes ineffective in treating the disease it was created to defeat. In this case, aromatase inhibitors are slowly becoming resistant to breast cancer cells, making it harder to find new drugs that can combat this trend. One way to combat it would be to understand the mechanisms that control this process. In one experiment, mice were treated with letrozole until their tumors became drug resistant. By doing so, they found that tyrosine kinase receptors, HER-2, adapter proteins (p-Shc and Grb-2), and all of the signaling proteins in the MAPK cascade were increased in tumors resistant to letrozole[21]. By knowing this, scientists can devise new drugs that can decrease the production of these proteins. Another way to overcome drug resistance for aromatase inhibitors would be to study patient pharmacogenomics. By screening patients to obtain their genomic file, physicians can have a better idea on which drug is best suited to the patient so that it avoids side effects and is effective[22]. As the knowledge

of aromatase inhibitors and the effects they produce in patients' bodies expand, there is still room for improvement to create better drugs for breast cancer treatment.

Methods

Overview

Mutant-type aromatase will be expressed and optimized in *E. coli* by adjusting induction time, incubation temperature, and cell strain from an already established protocol from the Lo lab[6]. Various inhibitors will also be screened using a CYP19 activity assay to identify potential inhibitors to test against aromatase. The long-term goal is to co-crystallize mutant aromatase with inhibitors of high potency.

Conventional Cloning

The mutant-type aromatase was produced via conventional cloning. This process consists of three main steps; creating the DNA insertion, inserting the DNA insert into a plasmid, and introducing the plasmid into *E. coli*. The aromatase insert, obtained from GenScript, was created by deleting 39 N-terminal amino acids from human aromatase, replacing them with 10 hydrophilic N-terminal residues, and adding 4 histidine residues at the C-terminus[6] (Figure 5).

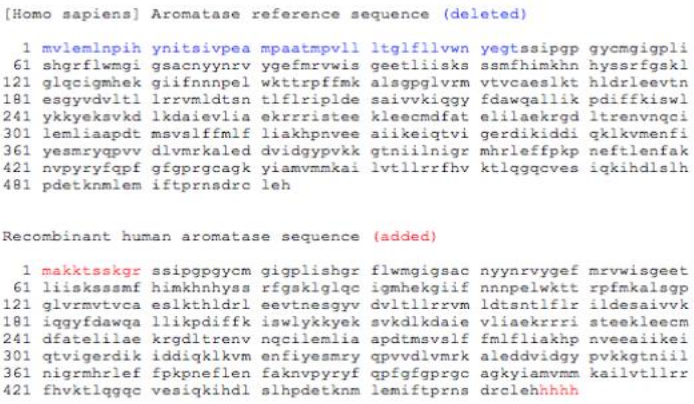

Figure 5: Sequence of original aromatase (above) and aromatase insert (below). The entire amino acid sequence of human aromatase with the 39 amino acids in blue represented the section to be deleted. For the recombinant aromatase, there was the addition of 10 hydrophilic amino acids on the N-terminus and 4 histidines on the C-terminus shown in red[6].

The 39 amino acids were deleted and replaced with 10 hydrophilic amino acids at the N-terminus to remove the predicted transmembrane domain, decreasing the likelihood of protein aggregation. The C-terminus had 4 histidine residues added to create a His-tag, making the protein easier to identify using Western blots and purify during the purification process. The cloning process was attempted many times in the lab with minimal success so it was decided to order the plasmid from GenScript to obtain the aromatase construct and proceed with the experiment. There were three attempts using a different vector, pCWori-A13AMO-aaCPRct, before a new plasmid was ordered from GenScript. The first attempt failed as the vector being used to clone the insert had an unknown size and sequence, making it hard to determine which construct was the correct one. The second attempt looked at dephosphorylating the vector to see if this would optimize ligation. Growth was still not seen on the plates. During the third attempt, an incorrect primer was discovered. However, the DNA concentration after PCR was still low. Since it was low, the construct could not be created as there was not enough of the insert to work with. The new plasmid construct contains the aromatase insert, replacing SacB in the pCW-LIC vector, which confers ampicillin resistance. Figure 6 shows the original plasmid.

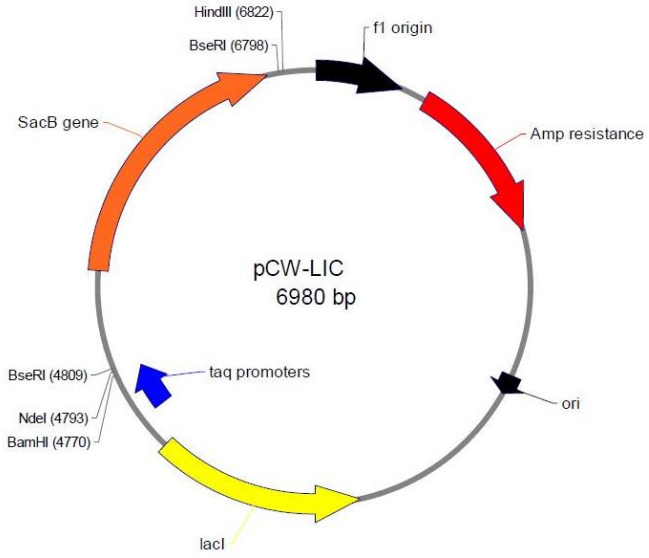

Figure 6: Diagram of pCW-LIC construct from GenBank. Construct contains 2 active taq promoters, an origin of replication (ORI), ampicillin (AMP) resistance, restriction sites, and a SacB gene which gets replaced with the aromatase insert.

Expression and Purification

The pCW-LIC expression plasmid, containing the aromatase insert, was used for cloning and transformed in E. cloni 10G competent cells. Rosetta2 *E. coli* cells were transformed with the aromatase construct and an expression vector containing HemH. HemH is the gene that encoded ferrochelatase, the enzyme that catalyzes the last step of heme biosynthesis[23]. The transformed cells were then plated on solid media mixed with antibiotics, carbenicillin and kanamycin. After overnight growth, one colony was isolated from the plate and resuspended in 2 mL of lysogeny broth (LB) with 2 µL 1,000X carbenicillin and kanamycin. After 16 hrs of growth overnight at 37°C with shaking, the cells were transferred into 100 mL of terrific broth (TB) and grown at 37°C with shaking at 250 rpm until the optimal density (OD) read between 0.5-0.8 at 600 nm. The temperature was dropped to 18°C and incubated for 1 hr. Following the incubation, 5 µM $FeCl_3$, 1 mM δ-ALA1, 1 mM IPTG and additional carbenicillin and kanomycin were added to the culture and grown for 48 hrs with continued shaking at 250 rpm. The cells were then harvested with a lysis buffer and centrifuged at 4,700 rpm for 10 min. The remaining pellet was washed 3 times using lysis buffer with 10X the pellet volume and stored in -80°C until later use. The pellet was resuspended in lysis buffer, lysed through sonication, and centrifuged at 20,000 g for 30 min at 4°C to attain the proteins soluble in the cell supernatant. Nickel affinity chromatography was used to separate mutant-type aromatase from its cellular debris with a wash and elution buffer. Aromatase has a molecular weight of 55 kDa[19]. SDS-PAGE and Western Blotting was used to confirm the presence of aromatase based on its size and presence of a His-tag.

Bioassays

The pure protein was characterized using an *in vitro* assay. The CYP19/MFC High Throughput Inhibitor Screening kit (Corning) was used to test aromatase activity. The kit provided native CYP19, 7-methoxy-4-trifluoromethyl coumarin (MFC), ketoconazole (KTZ), control protein, buffers, and cofactors. MFC is a fluorescent substrate that reacts with CYP19. KTZ is a positive control that inhibits CYP19 activity. The reaction involves CYP19 converting MFC into 7-hydroxy-4-trifluoromethyl coumarin (HFC). Since HFC is a fluorescent product, its activity can be measured by the intensity of the product. When an inhibitor such as KTZ is introduced into the system, it blocks the reactions from occurring; thus limiting product formation and the intensity of the fluorescence. Using this principle, multiple inhibitors were tested on the kit to determine its inhibitory concentration at 50% (IC_{50}) based on its fluorescence measured with a plate reader. Unfortunately, time constrains hindered the crystal screening portion of the experiment so crystals were not made. This would be the next step of the overall project.

Results and Discussion

Inhibition Assays

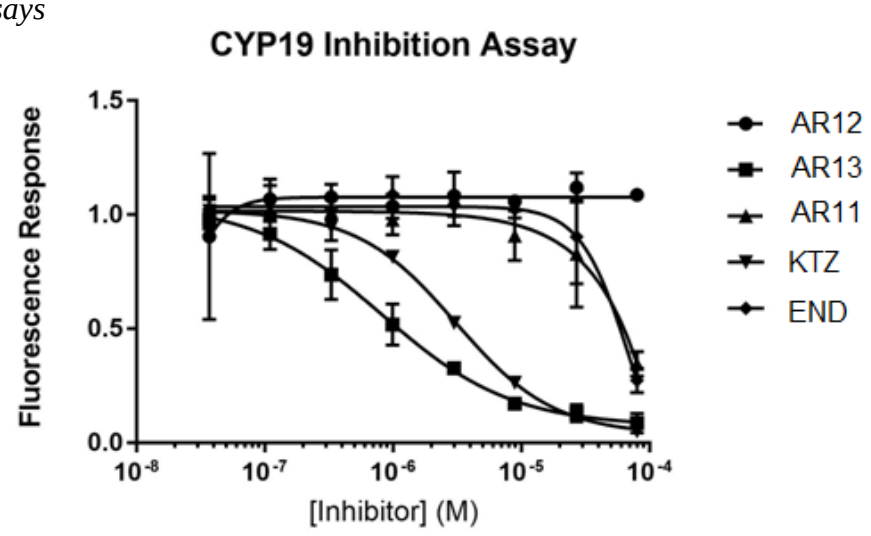

Figure 7: CYP19 Inhibition Assay with AR11, AR12, and AR13. AR11, AR12, and AR13were run against positive controls, ketoconazole (KTZ) and endoxifen (END), to test the effectiveness of the new inhibitors.

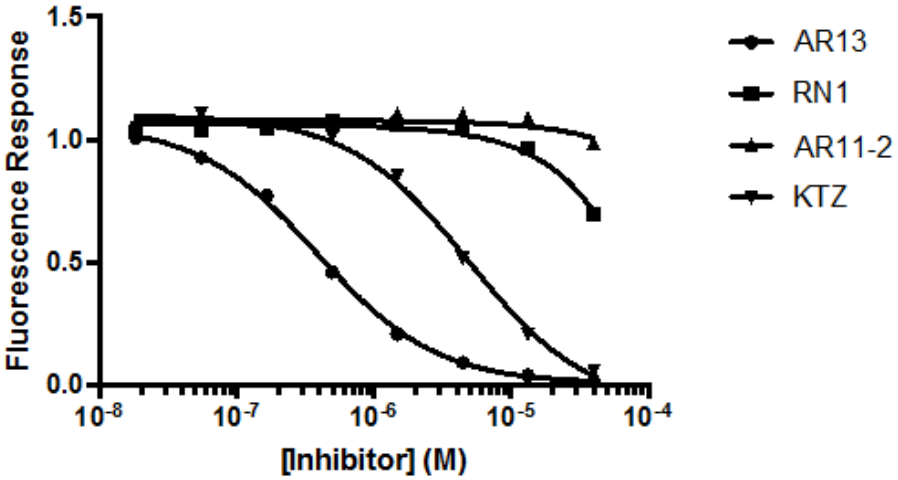

Figure 8: CYP19 Inhibition Assay with AR11 and AR13 analogs. AR11-2, an analog of AR11, and RN1, an AR13analog, were run against positive controls, KTZ and AR13, to test the effectiveness of the new inhibitors.

Three potential allosteric inhibitors, AR11, AR12, and AR13, were screened using the CYP19/MFC High Throughput Inhibitor Screening kit (Corning). They were run against ketoconazole (KTZ), a positive control from the kit, and endoxifen (END), a known aromatase inhibitor. A FluoDia T70 plate reader was used to measure aromatase activity from the fluorescent substrate 7-methoxy-4-trifluoromethyl coumarin (MFC) to 7-hydroxy-4-trifluoromethyl coumarin (HFC). The data was run in duplicate and averaged from two cycles. The inhibitors had a concentration of 25 μM in the first well. It is important to note that IC_{50} values are not concrete values as they change with varying conditions, especially the substrate affinity and whether the inhibitors are competitive or uncompetitive. IC_{50} values are only able to tell us the relative potencies of the various inhibitors being studied.

AR12 did not show significant inhibition. This suggests that it would not be a potent inhibitor as it would require a high dosage in order to be effective. AR11 produced an IC_{50} value of 25.35 μM, similar to END. AR13 produced an IC_{50} value of 0.41 μM, which was lower than KTZ. AR11 and AR13 had promising results as they performed at or better than the positive

controls that were being tested. The data also suggests that they would also require a normal or lower dosage, which is ideal when developing new drugs with reduced side effects.

Analogs of AR11 and AR13, AR11-2 and RN1, were then tested to determine if shape was the main contributor towards inhibitor potency. The IC_{50} value of AR11-2 was not obtained as it produced a large curve meaning this analog of AR11 was not potent. RN1 was not as potent as AR13 since it had a larger curve compared to AR13. It was determined that the inhibitor's binding affinities, not the structure, were causing the large changes in potency between the analogs. Once the allosteric inhibitors were chosen for future crystallization experiments, we focused on expression of mutant aromatase.

Co-Expression of CYP19/HemH in Rosetta2 Cells

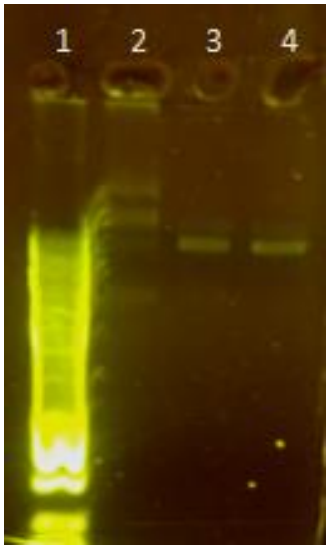

Figure 9: Gel Electrophoresis of Aromatase Plasmid Mini Prep. Lane 1: GeneRuler 1kb Ladder, Lane 2: Diluted Aromatase Construct from GenScript (120ng/mL), Lane 3 & 4: Human Aromatase (CYP19) Plasmid from Mini Prep DH5α cells.

To determine the presence of the gene insert in the plasmid, gel electrophoresis was run. Figure 9 showed the presence of 4 bands in the original diluted plasmid while only 2 distinct bands from the Mini Prep. However, the 2 bands corresponded with the bands present in the original construct so it was decided that the expression be continued as the plasmids obtained

from the Mini Prep were clean. More attention was paid towards the expression process as the difference in bands may imply improper protein expression and formation.

When aromatase (CYP19) and ferrochelatase (HemH) were co-expressed using two plasmids in Rosetta2 *E. coli* competent cells, the protein was not seen after Nickel affinity chromatography in Figure 10. The ladder was not helpful as it smeared from being in the first lane so only rough estimates could be made from the bands. It is important to note that Lane 5 is similar to Lane 4 because of residual protein being removed from the first wash. Lanes 8-9 also saw residual protein being removed from the first elution compared to Lane 7. There were some faint bands seen in Lanes 2-9 with Lanes 2-5 showing an additional band present in the lanes. Possible options for the different bands could be HemH (36 kDa monomer), ampicillin (AMP) (27 kDa), and chloramphenicol acetyltransferase (CAT) (26 kDa) as *E. coli* can express these enzymes from the plasmids construct they obtained.

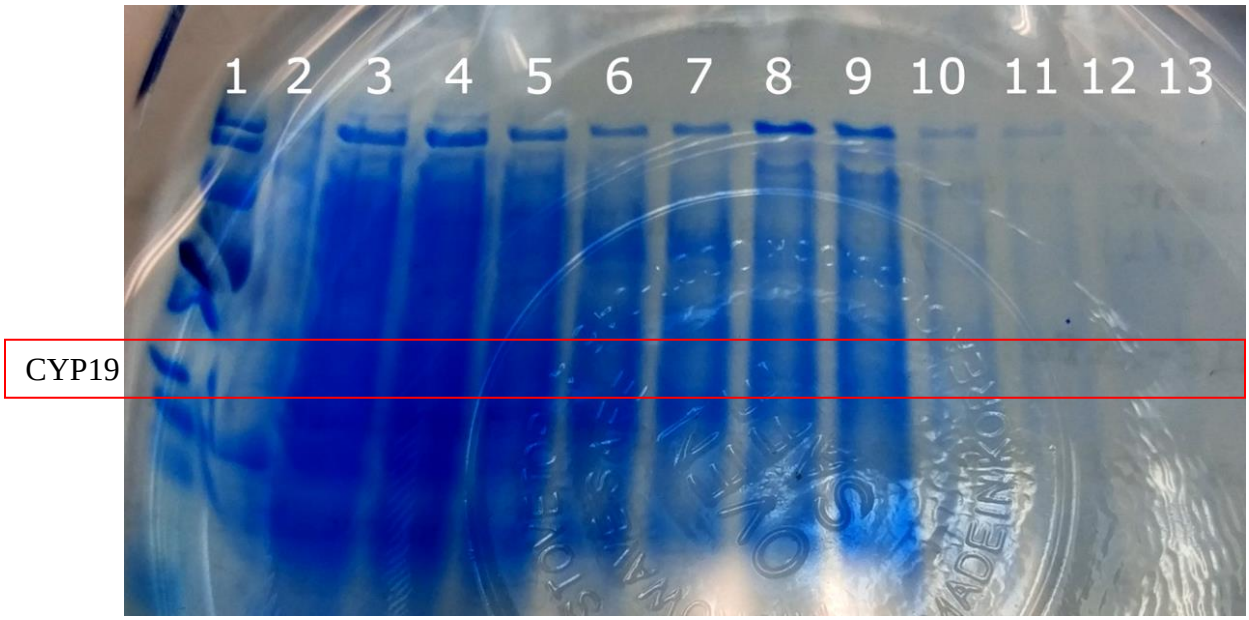

Figure 10: SDS-PAGE of Aromatase after His-Nickel Affinity Chromatography. Lane 1: PageRuler Prestained Protein Ladder, Lane 2: Pellet, Lane 3: Supernatant/Crude, Lane 4: Flow-through from chromatography, Lane 5-7: Washes from chromatography (3 times), Lane 8-13: Elution from chromatography (6 times). The co-expression of CYP19/HemH with Rosetta2 cells was used for this run.

A possibility for CYP19 not expressing could arise from the co-expression of two different plasmids. The HemH construct was ordered from GenScript. It originated from the pET vector and contains a T7 promoter, kanamycin (KAN) resistance, and an origin of replication (ORI). The CYP19 construct was also ordered from GenScript and originated from the pCWori^{+} vector. It contains a taq promoter, ampicillin (AMP) resistance, and an ORI. Looking at two constructs, one cause of discrepancy could arise from the ORI. Since the specific identity of the ORI is vague, it could be that the two plasmids have the same ORI. This could cause the bacteria to kick one of the plasmids out leaving the other plasmid behind for expression. It is also important to mention that the taq promoter is not as efficient as the T7 promoter as it is leaky suggesting expression of CYP19 may not be optimal.

Co-expression of CYP19 and HemH will be grown at the same time as the aromatase. If the CYP19 plasmid was disfavored during the co-expression, the two samples would show different protein results. If the plasmids were still present in the co-expression, then the two samples would have the same protein but be expressed at different concentrations. If the co-expression was not working properly, a new construct should be made where HemH and CYP19 are present in one plasmid to avoid the possibility of plasmid favoring.

Another reason CYP19 is not expressing could be from an improper sequence. Since the construct was made by the company Genscript, it was assumed that the sequence was correctly inserted inside the plasmid. However, nothing is confirmed unless the plasmid is actually sequenced. To test whether the correct sequence was inserted inside, the beginning and end of the CYP19 insert was sequenced and compared to the original template used. If there were errors at the ends, it was assumed construct was incorrectly made. If there were no mistakes, it meant that the construct was made correctly.

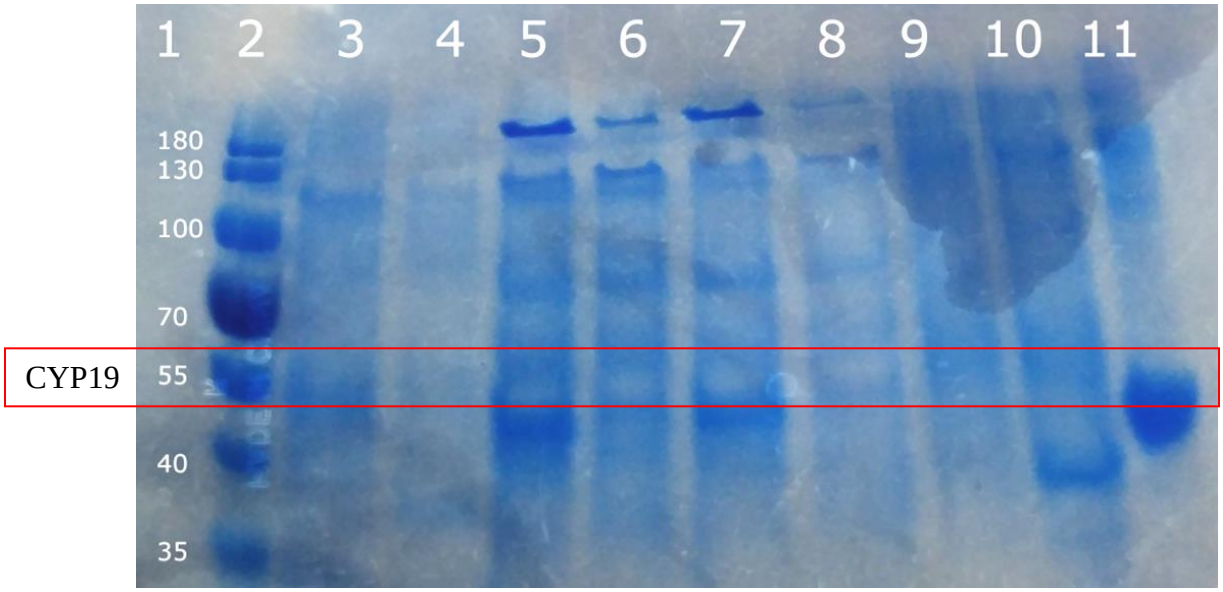

Figure 11: Second Run of Aromatase after His-Nickel Affinity Chromatography with SDS-PAGE. Lane 1: empty, Lane 2: PageRuler Prestained Protein Ladder, Lane 3: 2% SDS CYP19 pellet, Lane 4: 2% SDS CYP19/HemH pellet, Lane 5: Elution 1 CYP19, Lane 6: Elution 1 CYP19/HemH, Lane 7: Elution 2 CYP19, Lane 8: Elution 2 CYP19/HemH, Lane 9: CYP19 pellet with 4X loading dye, Lane 10: CYP19/HemH pellet with 4X loading dye, Lane 11: Ni-HRP Positive Control.

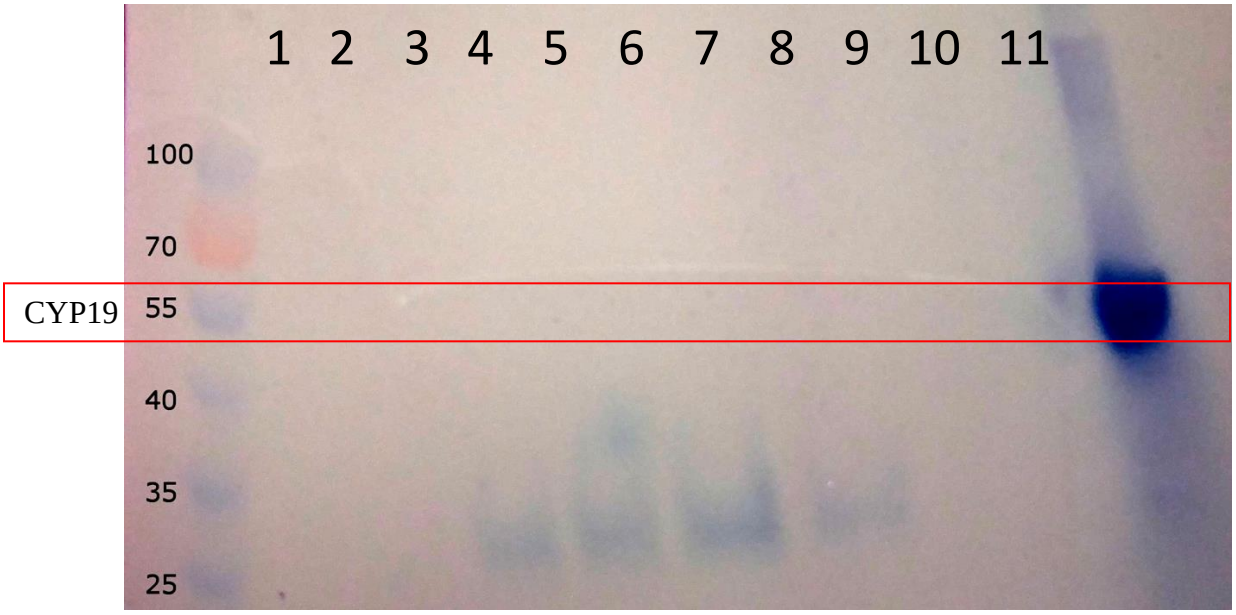

Figure 12: Second Run of Aromatase after His-Nickel Affinity Chromatography with Western Blot. Lane 1: empty, Lane 2: PageRuler Prestained Protein Ladder, Lane 3: 2% SDS CYP19 pellet, Lane 4: 2% SDS CYP19/HemH pellet, Lane 5: Elution 1 CYP19, Lane 6: Elution 1 CYP19/HemH, Lane 7: Elution 2 CYP19, Lane 8: Elution 2 CYP19/HemH, Lane 9: CYP19 pellet with 4X loading dye, Lane 10: CYP19/HemH pellet with 4X loading dye, Lane 11: Ni-HRP Positive Control.

Figure 11 and 12 showed the results of the purification from the second transformation of CYP19/HemH and CYP19 in Rosetta2 cells. CYP19 was present in the elutions of the SDS-

PAGE as there were bands at 55 kDa but are not visible in the Western blot. This suggests that the CYP19 HIS-tag is strong enough to be separated from other HIS-tagged proteins in the purification step but not strong enough to attach to the Nickel-Horseradish Peroxidase (Ni-HRP) probe used to visual the Western blot. The CYP19/HemH expression also produced less protein than the CYP19 expression as seen in the gel and varying pellets colors. CYP19/HemH produced a brown pellet while CYP19 produced a red pellet, the ideal color after iron incorporation in aromatase. CYP19 was not in the pellet after solublizing with 2% SDS and no significant bands were seen at 55 kDa.

The Western reaffirmed that other HIS-tagged proteins are present in the expression as there were visible bands above 25 kDa. The sequencing data of CYP19 plasmid and CYP19 Mini Prep were also obtained. The forward sequences were successful while the reverse sequences were not. Since the reverse primers were causing issues, a different primer was used to assure the HIS-tag was present in the construct as the successful forward sequence confirmed CYP19 was inserted into pCW-LIC. A forward primer that bound to the nucleotide 1,230 of the insert was used to bind near the end to confirm the presence and length of the HIS-tag attached to CYP19. From the data collected so far, it was decided to test time points of the expression with retransformed CYP19 cells as it seemed to produce more protein than the CYP19/HemH co-expression.

CYP19 in Rosetta2 cells were grown under normal conditions and 100 mL samples were taken starting from 18 hrs. This continued every 6 hrs instead of harvesting the cells after 48 hr to observe the varying amounts of protein being produced at each time point. However, the results of the time trials were inconclusive as the samples were found not shaking until 18 hrs after initial growth. Because of this, it was decided to toss all samples except 48 hrs in order to

perform a tryptic digest. This was done to determine whether there was functional protein present in the culture, even without the required shaking. Unfortunately, the tryptic digest did not work so the data was not very informative. The sequencing data told us that the His-tag was inserted correctly as the 4 histidine residues were seen in the sequence data. A different antibody should be used to detect 4 histidine residues specifically as the current probe can only detect 6 histidine residues. Since the time trials were uninformative, the next variable to focus on was growth temperature. Two cultures were grown normally with the only difference being temperature. One flask would be grown at 28°C and while the other grown at 18°C.

Temperature Modifications for CYP19 Expression in Rosetta2 Cells

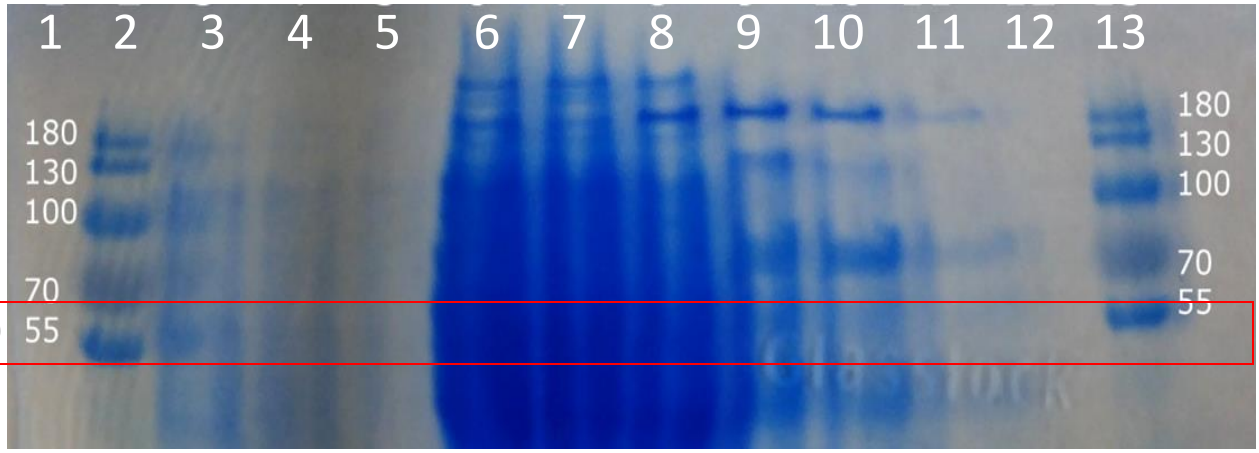

Figure 13: SDS-PAGE of Aromatase Temperature Trials after His-Nickel Affinity Chromatography. Lane 1: empty, Lane 2: PageRuler Prestained Protein Ladder, Lane 3-5: Elutions from cells grown at 18°C (3 times), Lane 6: Supernatant/Crude from cells grown at 28°C, Lane 7: Flow-through from cells grown at 28°C, Lane 8: Wash from cells grown at 28°C, Lane 9-11: Elutions from cell grown at 28°C (3 times), Lane 12: empty, Lane 13: PageRuler Prestained Protein Ladder.

Figure 13 depicts elevated expression levels at 28°C over 18°C. This was indicated by increased intensity of the gel band. Furthermore, the mass of the cell pellet after harvest was greater and it exhibited a deeper red color. Only part of the purification from the 18°C growth was run on a Western blot while the 28°C growth was examined in more detailed. From the Western blot, contaminants were seen at 180 kDa since there was a strong band at that molecular weight. A possible contaminant at 180 kDa could be RNase E as it falls in the same size and is

present in bacterial expression. There also seemed to be faint bands present in the elutions at 55 kDa suggesting the production of CYP19. However, 55 kDa should be the strongest band and not the contaminant at 180 kDa. A new purification method was tested since the His-tag bound poorly to the Ni-HRP used to visualize the Western blot. It was assumed that the His-tag on CYP19 might not attach to the Nickel affinity chromatography correctly since it had a weak affinity for 4 histidine residues, causing some of the protein to be lost in the wash and flow-through. The reason for preferential growth at 28°C comes from a higher optimal growth density at higher temperatures than at lower temperatures. While growth at a lower temperature produces more functional protein as it insures slow peptide folding and incorporation, it does inhibit some protein production. To fully confirm the production of CYP19, 2 vials of cells would be grown at 28°C for 48 hrs with the only difference being the presence or absence of IPTG to induce the cells.

Induction Tests for CYP19 Expression in DH5α Cells

When the cells were grown in the following conditions, there was no protein present in either sample. It is interesting to note that the uninduced cells had the correct supernatant color of a red hue even though the pellet was much larger and chunkier than the induced cells. The reason could arise from the presence of iron in the media as it may not have been incorporated properly into the protein. After concentrating both samples, the gel seemed to be inconclusive as there were lots of contaminants present but CYP19 was nowhere to be found. It was decided to reattempt the co-expression of CYP19 and HemH in DH5α *E. coli* competent cells and perform colony-PCR to test whether the samples being grown had the insert present.

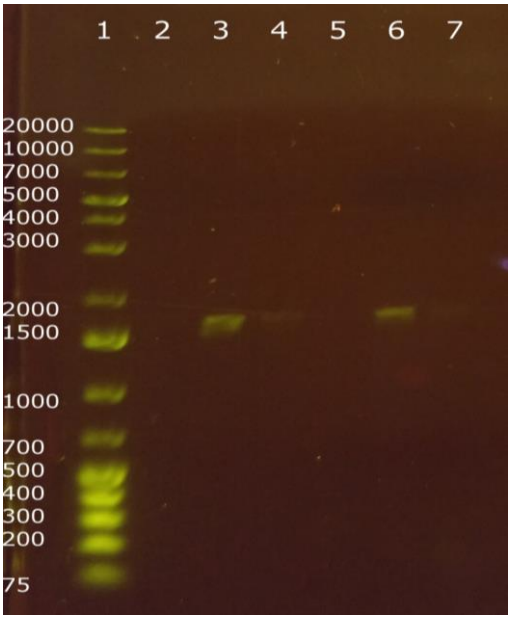

Figure 14: Gel Electrophoresis of Colony PCR with CYP19 and CYP19/HemH Colonies in DH5α. Lane 1: GeneRuler 1kb Ladder, Lane 2-4: Three different colonies from CYP19 plate (A1, B1, C1), Lane 5-7: Three different colonies from CYP19/HemH plate (A2, B2, C2).

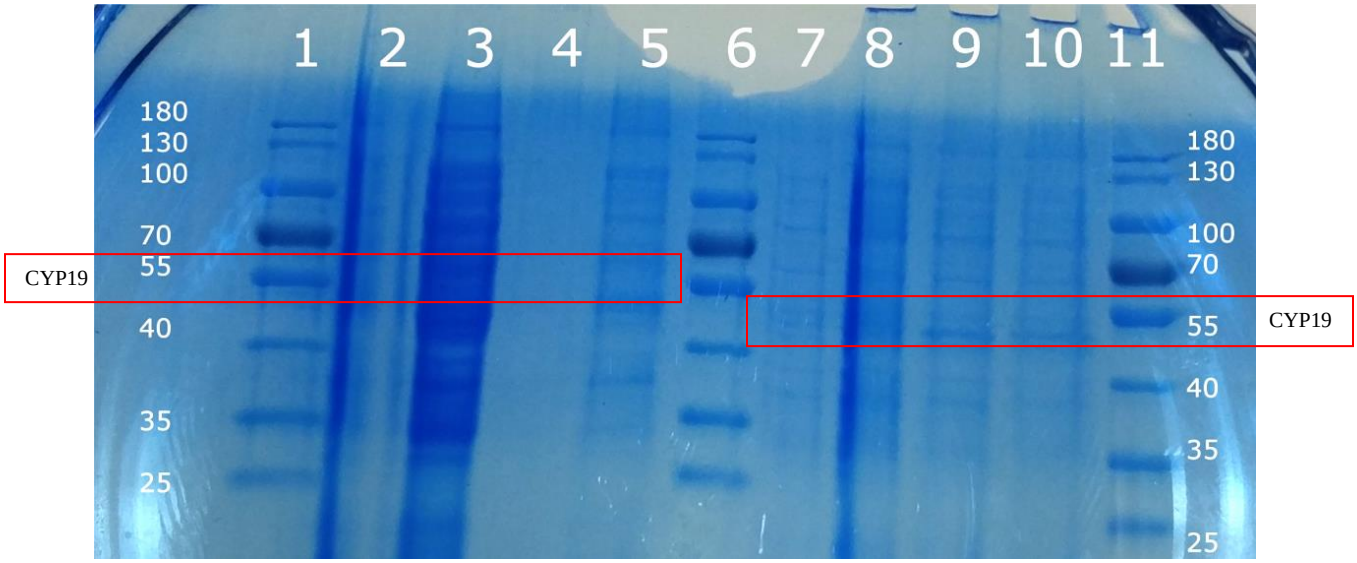

Figure 15: SDS-PAGE of CYP19 and CYP19/HemH alongside CYP19 with and without IPTG induction in DH5α. Lane 1: PageRuler Prestained Protein Ladder, Lane 2: CYP19 pellet, Lane 3: CYP19 supernatant, Lane 4: CYP19/HemH pellet, Lane 5: CYP19/HemH supernatant, Lane 6: PageRuler Prestained Protein Ladder, Lane 7: CYP19 with IPTG pellet, Lane 8: CYP19 pellet without IPTG, Lane 9: CYP19 with IPTG supernatant, Lane 10: CYP19 without IPTG supernatant, Lane 11: PageRuler Prestained Protein Ladder.

A new plate of CYP19 and CYP19/HemH was expressed on their respective antibiotic plates. After incubating, 3 colonies from each plate were picked and used to run colony-PCR. Of the 6 samples, only 2 colonies had inserts at 1.5 kb. This was concerning as all of the samples were expected to contain the insert but the experiment was continued and the issue noted. Using

the 2 colonies, CYP19 and CYP19/HemH were grown in 50 mL of TB and washed with phosphate buffer, pH 7.4. The cells were also solubilized with 1 mg/mL lysosyme and 1% Tween20 to help lyse the cells open. The pellet and supernatant were analyzed using SDS-PAGE and Western blot. Ni-HRP and Anti-Aromatase antibodies were tested on the Western blots to determine the presence of CYP19 in the supernatant. Since the Western blots failed to show any bands, it was not included in the data. It was also decided that the Anti-Aromatase antibody would be for future Western blot development as it would be more CYP19-specific compared to the Ni-HRP.

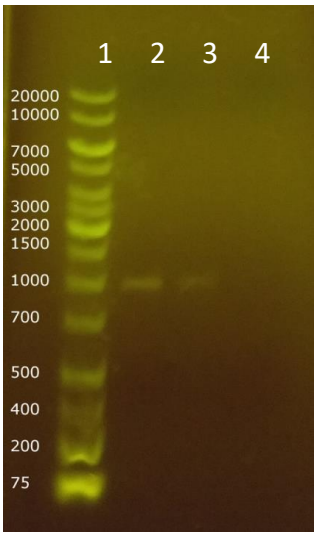

Figure 16: Gel Electrophoresis of Colony PCR with CYP19 in DH5*α*. Lane 1: GeneRuler 1kb Ladder, Lane 2-4: Three different colonies from CYP19 plate (A, B, C).

Colony-PCR was repeated on CYP19 as all of colony B1 from the previous experiment was used up before being plated. The results were more worrisome as the bands were at 1 kb instead of 1.5 kb. However, since it was the only band present and a 1.5 kb band was present in the previous PCR, the experiment continued on. It is also important to note that it could have been the use of a different ladder that was giving different band sizes. Since there was still no progress in expression and visualization of CYP19, it was decided to induce the cells again in the presence and absence of IPTG but maintain the temperature at 37°C.

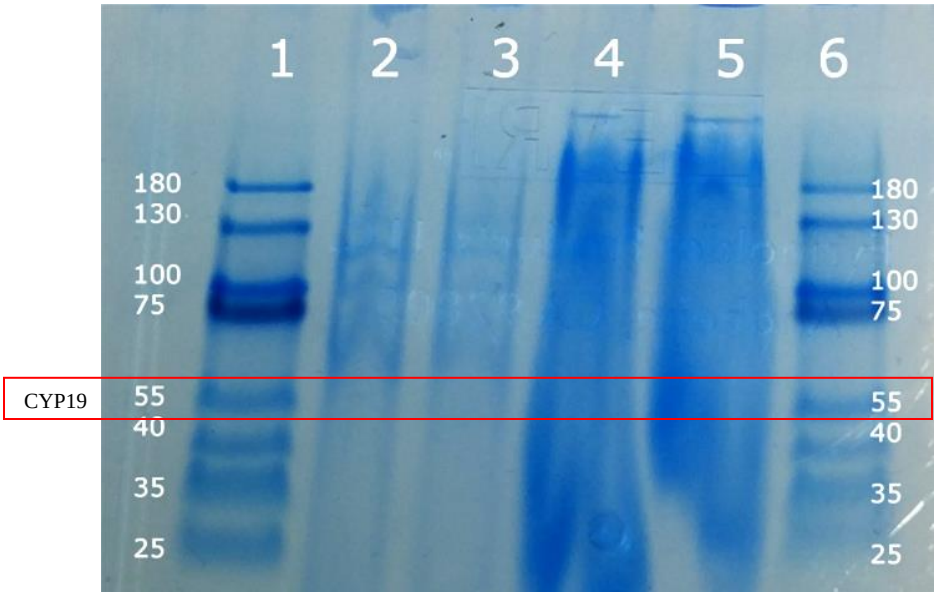

Figure 17: SDS-PAGE of CYP19 with and without IPTG induction in DH5α. Lane 1: PageRuler Prestained Protein Ladder, Lane 2: CYP19 pellet with IPTG, Lane 3: CYP19 pellet without IPTG, Lane 4: CYP19 supernatant with IPTG, Lane 5: CYP19 supernatant without IPTG, Lane 6: PageRuler Prestained Protein Ladder.

The reason for maintaining the induction at a high temperature was to see if CYP19 was being produced. Even if inclusion bodies and inactive protein formed, the goal was to see if anything was being made. If nothing was found in the SDS-PAGE and Western blot, it could be assumed that the promoter was not functioning properly. This meant that a new construct of CYP19 should be made and the experiments repeated. The pellet and supernatant of this experiment were run on the same gel (Figure 17) with the previous experiment so a side-by-side comparison could be done. The SDS-PAGE and Western blot for all of the samples being tested did not show any signs of aromatase. However, there was a large band around 15 kDa on the SDS-PAGE. This is interesting as there was nothing seen in the Western blot at the size. Another interesting note is that the cells induced with IPTG looked exactly like the cells without IPTG induction in terms of protein expression. This suggests that the promoter may not be functioning properly so a sequence was done on the construct to see if this was the case.

Basal Expression Levels for CYP19 in Rosetta2 Cells

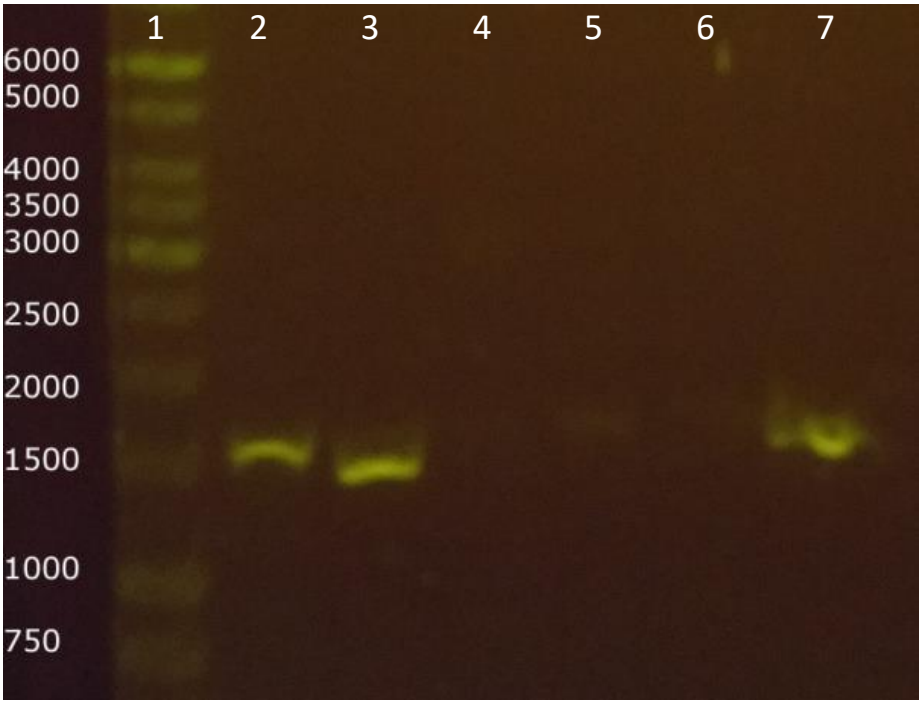

Figure 18: Gel Electrophoresis of Colony PCR with CYP19 and pCW-LIC in Rosetta2. Lane 1: GeneRuler 1kb Ladder, Lane 2-4: Three different colonies from CYP19 plate (A, B, C), Lane 5-7: Three different colonies from pCW-LIC plate (A, B, C).

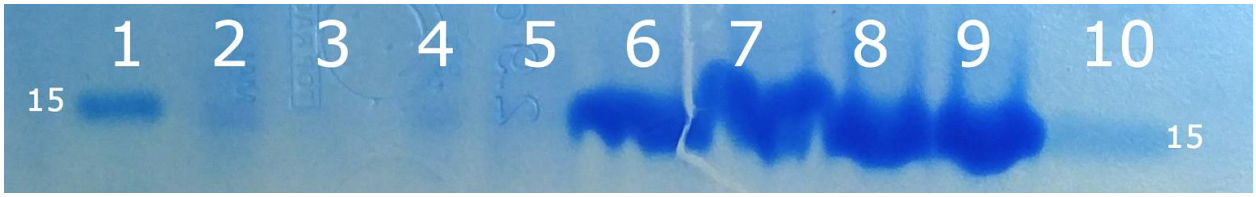

Figure 19: SDS-PAGE with CYP19 and pCW-LIC in Rosetta2. Lane 1: PageRuler Prestained Protein Ladder, Lane 2: CYP19 with IPTG pellet, Lane 3: CYP19 without IPTG pellet, Lane 4: pCW-LIC (B) pellet, Lane 5: pCW-LIC (C) pellet, Lane 6: CYP19 with IPTG supernatant, Lane 7: CYP19 without IPTG supernatant, Lane 8: pCW-LIC (B) supernatant, Lane 9: pCW-LIC (C) supernatant, Lane 10: PageRuler Prestained Protein Ladder.

The final step was to compare CYP19 to the empty pCW-LIC vector in Rosetta2 cells to confirm the construct was not working properly. The reason for the change in cell strain is because DH5α is used primarily for cloning instead of expression. Since Rosetta2 is a strain used specifically for expression, it could tell us whether cell strain was the issue or if the vector itself was defective. The cells were transformed and colony-PCR was run to ensure the transformation worked. It was expected that bands would be present in CYP19 and not the vector as it should not contain the insert. However, the gel seemed to show a strong band in Lane 7 and faint bands

in Lane 5-6 for pCW-LIC. Because of this, two negative controls (colony B and C from pCW-LIC) were expressed alongside colony B from CYP19 and induced at 37°C to confirm CYP19 was being produced in the cells.

Observing Figure 19, all of the samples looked identical. There was also no band at 55 kDa but a large band at 15 kDa suggesting the presence of contaminants. This protein could have been chaperonin 10 (GroES) as the high induction temperatures may have caused the protein to shift down the gel. GroES is an important protein that works in conjunction with chaperonin 60 (GroEL) for proper protein folding. Since GroEL was not seen in the gel but GroES was, it could be that the bacteria were not producing it. If the bacteria were only producing one of the two major protein chaperones, it could be that CYP19 was being made but not being folded correctly. Examining the sequencing results of the promoter, it was determined to be mutation-free. It was decided that a new vector should be constructed and the experiments performed would be done again.

Conclusions

Overall, novel allosteric inhibitors were discovered and expression conditions for aromatase tested. AR11 and AR13 showed promising results as possible aromatase inhibitors but needed further studies to confirm this. Co-expression of CYP19 and HemH did not show improved expression of CYP19. Increased incubation time showed that longer induction time increased protein expression though it could be from increased cell density. Increases in induction temperature showed increased protein expression though this could also be from increased cell density. Changes in cell strain showed no difference in protein expression so Rosetta2 competent cells were still used as they were expression-specific. Induction studies

showed no difference in protein expression either, suggesting a mutation in the promoter region which could affect CYP19 expression. However, the promoter showed no signs of mutations suggesting a different issue in cloning occurred. For future studies, it is hoped that aromatase will be expressed and purified in ample amounts using a new construct. More assays should also be run to determine the modality and K_I of the newfound inhibitors. Finally, co-crystallization of mutant aromatase to said inhibitors should be created via crystal screens and their crystal structure determined so more aromatase structures can be obtained.

Bibliography

1. Breastcancer.org. U.S. Breast Cancer Statistics. *Breastcancer.org* (2016). Available at: http://www.breastcancer.org/symptoms/understand_bc/statistics. (Accessed: 15th February 2016)
2. Chumsri, S., Howes, T., Bao, T., Sabnis, G. & Brodie, A. Aromatase, aromatase inhibitors, and breast cancer. *J. Steroid Biochem. Mol. Biol.* **125,** 13–22 (2011).
3. Brueggemeier, R. W., Hackett, J. C. & Diaz-Cruz, E. S. Aromatase Inhibitors in the Treatment of Breast Cancer. *Endocr. Rev.* **26,** 331–345 (2005).
4. M. T. Zivian & B. Salgado. SIDE EFFECTS REVISITED: Women's Experiences With Aromatase Inhibitors A Report From Breast Cancer Action. (2008). Available at: http://archive.bcaction.org/uploads/PDF/AIReport.pdf. (Accessed: 24th June 2016)
5. Bhatnagar, A. S. The discovery and mechanism of action of letrozole. *Breast Cancer Res. Treat.* **105,** 7–17 (2007).
6. Lo, J. *et al.* Structural Basis for the Functional Roles of Critical Residues in Human Cytochrome P450 Aromatase. *Biochemistry (Mosc.)* **52,** 5821–5829 (2013).

7. Brueggemeier, R. W., Richards, J. A., Joomprabutra, S., Bhat, A. S. & Whetstone, J. L. Molecular pharmacology of aromatase and its regulation by endogenous and exogenous agents. *J. Steroid Biochem. Mol. Biol.* **79,** 75–84 (2001).
8. Vaz, A. in *Drug metabolizing enzymes: cytochrome P450 and other enzymes in drug discovery and development* (FontisMedia SA, 2003).
9. Yoshimoto, F. K. & Guengerich, F. P. Mechanism of the Third Oxidative Step in the Conversion of Androgens to Estrogens by Cytochrome P450 19A1 Steroid Aromatase. *J. Am. Chem. Soc.* **136,** 15016–15025 (2014).
10. Sohl, C. D. & Guengerich, F. P. Kinetic Analysis of the Three-step Steroid Aromatase Reaction of Human Cytochrome P450 19A1. *J. Biol. Chem.* **285,** 17734–17743 (2010).
11. Yue, W. *et al.* The potential role of estrogen in aromatase regulation in the breast. *J. Steroid Biochem. Mol. Biol.* **79,** 157–164 (2001).
12. Chen, Y. *et al.* Combined Src and Aromatase Inhibition Impairs Human Breast Cancer Growth In vivo and Bypass Pathways Are Activated in AZD0530-Resistant Tumors. *Am. Assoc. Cancer Res.* **15,** 3396–3405 (2009).
13. Rodriguez-Antona, C. & Ingelman-Sundberg, M. Cytochrome P450 pharmacogenetics and cancer. *Oncogene* **25,** 1679–1691 (2006).
14. Ghosh, D., Griswold, J., Erman, M. & Pangborn, W. Structural basis for androgen specificity and oestrogen synthesis in human aromatase. *Nature* **457,** 219–223 (2009).
15. Kagawa, N., Hori, H., Waterman, M. R. & Yoshioka, S. Characterization of stable human aromatase expressed in E. coli. *Steroids* **69,** 235–243 (2004).
16. Kagawa, N. Efficient expression of human aromatase (CYP19) in E. coli. *Methods Mol. Biol. Clifton NJ* **705,** 109–122 (2011).

17. Kim, D.-H. *et al.* Heterologous expression and characterization of wild-type human cytochrome P450 1A2 without conventional N-terminal modification in Escherichia coli. *Protein Expr. Purif.* **57,** 188–200 (2008).
18. Hobler, A. *et al.* Human aldosterone synthase: Recombinant expression in E. coli and purification enables a detailed biochemical analysis of the protein on the molecular level. *J. Steroid Biochem. Mol. Biol.* **132,** 57–65 (2012).
19. Biegon, A., Alia-Klein, N. & Fowler, J. S. Potential contribution of aromatase inhibition to the effects of nicotine and related compounds on the brain. *Neuropharmacology* **3,** 185 (2012).
20. Shimodaira, M. & Nakayama, T. in *Cancer Etiology, Diagnosis and Treatments : Aromatase Inhibitors : Types, Mode of Action and Indications* 1–20 (Nova Science Publishers, Inc., 2009).
21. Jordan, V. C. & Brodie, A. M. H. Development and Evolution of Therapies Targeted to the Estrogen Receptor for the Treatment and Prevention of Breast Cancer. *Steroids* **72,** 7–25 (2007).
22. Coley, H. M. Mechanisms and strategies to overcome chemotherapy resistance in metastatic breast cancer. *Cancer Treat. Rev.* **34,** 378–390 (2008).
23. Sudhamsu, J. *et al.* Co-expression of ferrochelatase allows for complete heme incorporation into recombinant proteins produced in E. coli. *Protein Expr. Purif.* **73,** 78–82 (2010).